Print

Lesson 13: Hypothesis Testing - Study Notes

Slide 1:

Hypothesis Testing

We make decisions every day. Whether it is about what to have for breakfast, whether or not to cross a street, or what to watch on television, our everyday decisions range from very important to insignificant. Regardless of the potential impact, we shall arrive at our conclusion in the same way. First, we identify the problem and weigh the alternatives. Then, based on past experience, personal beliefs, and/or concrete evidence, we arrive at a decision and take action. Hypothesis testing follows a very similar pattern. However, instead of using personal beliefs and experiences, hypothesis testing uses sample statistics to arrive at its conclusion. For instance, Canadian pollsters study small samples of people in order to make predictions about how an entire country will vote at election time. Similarly, we may wish to compare the anxiety of male and female graduate students in Canada when it comes to math, or the effect of a particular drug on children with Attention Deficit Disorder.

While confidence intervals were used to estimate the value of a parameter, hypothesis testing is a way of deciding whether a statement concerning a parameter is true or false; that is, we must test a hypothesis about a parameter. As we noticed with confidence intervals, the fact that we are using samples instead of the population as a whole introduces some error in our estimate. So, we cannot base our conclusion with regards to the hypothesis solely on discrepancies between the sample statistic and the population parameter. Instead, we look for significant differences (or lack thereof) and incorporate probabilities to aid in our decision making.

Slide 2:

Example

To illustrate the general concepts involved in this kind of decision problem, suppose that an ambulance company wants to test a hospital's claim that the average time it takes patients to make it to the emergency ward is 20 minutes. From a random sample of 40 patients, the company finds that the mean time is 18.5 minutes with a standard deviation of 4. Please test the hospital's claim at a level of significance of 0.05.

The important values that we need to retain from the infomation provided are the following:

- 20 minutes (population mean)
- 40 patients (sample size)
- 18.5 minutes (sample mean; the mean based on the 40 patients sampled)
- sample standard deviation = 4
- significance level = 0.05 (can also be presented as 95% confidence level)

Now let's solve this example in a step-by-step process.

Slide 3:

Step 1: The Null and Alternate Hypothesis

When preparing to tackle a hypothesis test, the first step is to set up your hypotheses. Each test will have two hypotheses: the null hypothesis and the alternate hypothesis.

The **null hypothesis** (Ho) can be loosely described as a statement identifying the population parameter as being a specific value. It may also be used to deny any relationship, effect, or change caused by an independent variable (a treatment) on the population of interest. Here are some examples:

a) Ho: The sky is blue.
b) Ho: The drug has no effect on the patients.
c) Ho: The population mean (μ) is equal to 55.

The **alternate (or research) hypothesis** (Ha) is basically a statement that complements the null hypothesis. Since the two hypotheses are mutually exclusive, Ha is the opposite of Ho, so that the rejection of the null hypothesis would mean that the alternate hypothesis must be true. For example:

a) Ha: The sky is not blue.
b) Ha: The drug has an effect on the patients.
c) Ha: The population mean (μ) is not equal to 55.

In the Cardiack case, our hypotheses would look like this:

style="FONT-SIZE: medium">Ho: $\mu = 20$ minutes
Ha: $\mu \neq$ style="FONT-SIZE: medium"> 20 minutes

This particular type of hypothesis sets the table for a **two-sided test** because we are only interested in determining if a difference exists. We do not care if it is more or less than 20 minutes; we are only concerned with identifying some sort of significant difference...if it exists. But what if we altered the claim?

Slide 4:

Step 1: The Null and Alternate Hypothesis (Cont'd)

Suppose that the same ambulance company wanted to test a hospital's claim that their newest ambulance service, NotSoPatient, has an average travel time of **less than** 20 minutes. Here, we are interested in values being less than 20 (< 20) if the claim is to be supported. What must we do to the hypotheses to reflect this? We must alter one of the hypotheses to $\mu < 20$ minutes, and, consequently, the other to $\mu \geq 20$ minutes (since it must be the opposite).

As a general rule, **anything with an "=" sign (including ≥ and ≤) goes into Ho**. Therefore, the new hypotheses will look like this:

style="FONT-SIZE: medium">Ho: $\mu \geq 20$ minutes
Ha: $\mu < 20$ minutes

This is a **one-sided test** since we are only interested in values that are significantly less than 20.

Notice that the claim (< 20) is now represented by Ha, whereas in our first example, it was in Ho (= 20). Therefore, the claim is not what dictates what goes into Ho and Ha, but rather the mathematical symbol (<, =, ≥, etc.) associated with the claim.

Slide 5:

Step 1: The Null and Alternate Hypothesis (Cont'd)

The following is a summary of the three possible scenarios for the null and alternate hypotheses:

>Scenario 1	>Scenario 2	>Scenario 3
style="FONT-SIZE: medium">Ho: $\mu = \mu_o$	style="FONT-SIZE: medium">Ho: $\mu \geq \mu_o$	style="FONT-SIZE: medium">Ho: $\mu \leq \mu_o$
style="FONT-SIZE: medium">Ha: $\mu \neq \mu_o$	style="FONT-SIZE: medium">Ha: $\mu < \mu_o$	style="FONT-SIZE: medium">Ha: $\mu > \mu_o$

Where μ_o is the proposed parameter value

Common statements and their corresponding mathematical notations include, but are not limited to:

=	≠	≤	>	≥	<
Is	Is not	Is at most	Is better	Is at least	Is less
Same as	Not same as	Is no greater than	Is greater than	Is not less than	Is worse than
Not different	Is different	Is no more than	Is superior to	Is no worse than	Is inferior to

Slide 6:

Step 1: The Null and Alternate Hypothesis (Cont'd)

Example 1

Please determine the null and alternate hypotheses of the following claims:

- The motorists do indeed respect the 50 km/h speed limit.
- The average GPA at Concordia University is 2.4.
- The Montreal Canadiens have averaged no more than 220 goals over the last decade.

- The new nasal spray is much more effective than the usual 2 hours of relief that the other brands provide.
- The new technique produces at least 30 items a minute.

Answer:

Determine Ho and Ha.

- Ho: $\mu \leq 50$, Ha: $\mu > 50$
- Ho: $\mu = 2.4$, Ha: $\mu \neq 2.4$
- Ho: $\mu \leq 220$, Ha: $\mu > 220$
- Ho: $\mu \leq 2$, Ha: $\mu > 2$
- Ho: $\mu \geq 30$, Ha: $\mu < 30$

Slide 7:

Step 2: Determine the Significance Level

The level of significance, denoted by alpha (α), provides us with the means to distinguish between a difference due to chance, and a difference that is significantly due to more than chance. For example, if the confidence level has been set at 90%, then there is a 10% chance that the result obtained is significant enough to reject the null hypothesis. We will look at this in more detail later on in this section.

If the significance level is not assigned in the problem, then you have the enviable task of choosing one. The most common alpha levels in scientific experiments are 0.01 (99%), 0.02 (98%), 0.05 (95%) and 0.1 (90%).

In our case (the sample problem), we are told to use a level of significance of 0.05.

Slide 8:

Step 2: Determine the Significance Level (Cont'd)

Selecting a Method:

There are three procedures to complete a hypothesis test:

- The Classical Approach (Graphical)
- The Confidence Interval
- The Probability-Value Approach

The **classical** approach, also known as the **graphical method**, involves the comparison of a test statistic to a fictional boundary on the normal distribution (or t-distribution). This boundary acts as a dividing point between the area represented by α (the **critical region**) and the area represented by 1 -α. If the test statistic falls in the region represented by α, then we would have ample evidence to reject the null hypothesis, and accept Ha.

The **confidence interval** follows the same steps we investigated earlier in this lesson, but is applied to determining the probable value of >μ. Basically, we construct a confidence interval to see if the postulated population mean falls within the upper and lower limits. If μ is not contained in the interval, then Ho is rejected in favour of Ha. Unfortunately, **this technique is limited to two-sided tests**.

The **probability-value** approach, or p-value method, focuses on the probability that the test statistic could be the value it is, or a more extreme value (in the direction of the alternate hypothesis). It uses the significance value as a standardised bar such that if the probability of obtaining a given test statistic is less than the alpha value, we have enough evidence to reject the null hypothesis and go with Ha.

For beginners, it is suggested that you employ a combination of the classical approach and the p-value method. What this does is perform two tests at the same time that should give the same end result. If they do not, then chances are that you have made an error somewhere.

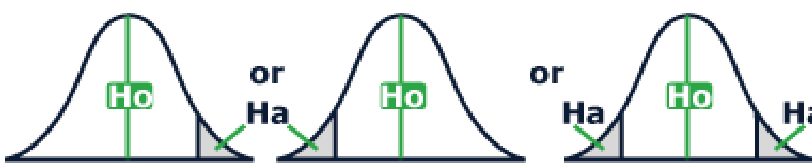
For the ongoing example, we will choose the classical approach and do a quick p-value to double-check our final answer.

Slide 9:

Step 3: Critical Region

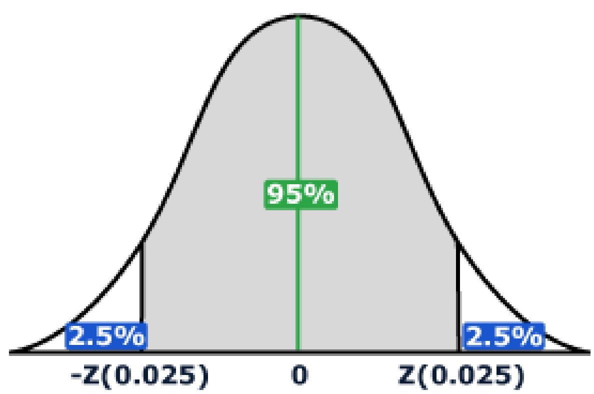
The critical region is the name we give to the small area of the distribution that represents the alternative hypothesis (Ha). The exact area of the critical region is determined by alpha (α), the significance level. For instance, if the confidence level was set at 98%, then the critical region's area is 0.02.

Furthermore, the location of the critical region on the distribution (extreme right, extreme left, or both extremes) is determined by Ha.

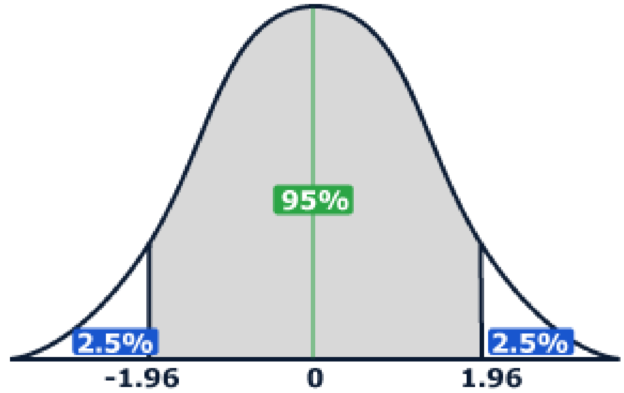


We are interested in finding the specific value that will serve as our boundary for the critical region. This value just happens to be the **confidence coefficient**.

In our case, the Ha of "not equal to" means that the critical region will be found at both extremes of the distribution. Therefore, the alpha value of 0.05 (since our test is at the 95% confidence level) will be split in half. The critical area on each side will be 0.025. Since the sample size is large (n = 40), we will use the standard normal distribution to find the exact values that contain 95%, or an area of 0.95, between them.



By looking at the table (or by remembering a previous example), we found the **critical values** to be +1.96 and -1.96. A critical value refers to the boundary separating the area representing the null hypothesis, and the critical region (Ha).



Slide 10:

Step 4: Test Statistic

The **test statistic** refers to the random variable that is calculated from the sample data, and is used to make a decision with regards to the null hypothesis. The purpose of the test statistic is to determine whether or not the data from the sample differs from the parameter because of chance, or because it is significantly different.

The test statistic, denoted by a "*" (z* for statistics from the standard normal distribution, t* if we employed the t-distribution), uses a very familiar formula:

$$test\ statistic = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

Where "σ" can be estimated by "s" if it is unknown

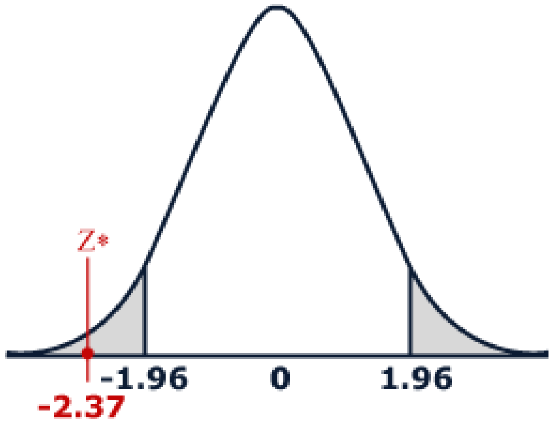
In our example, the test statistic is described as z* since we are using the standard normal distribution. By substituting the sample statistics into the test statistic equation, we get:

$$Z^* = \frac{18.5-20}{4/\sqrt{40}} = -2.37$$

Slide 11:

Step 5: Drawing Conclusions

The test statistic is now compared to the critical region.



As you can see, the test statistic falls in the critical region in the extreme left tail of the distribution. This is enough evidence to reject the null hypothesis (Ho) and accept the alternate one (Ha). In other words, there is enough evidence to reject the hospital's claim that the time of arrival to the hospital is 20 minutes.

Decision Summary

>The test statistic...	>Does not fall in the critical region	>Falls within the critical region
>Ho	Is not rejected	Is rejected
>Ha	Is not accepted	Is accepted

It is important to refer back to the original claim when writing your conclusion. The fact that we have rejected Ho does not mean anything on its own. We must refer back to step 1 and see what Ho and Ha represent, and mention this in our final statement.

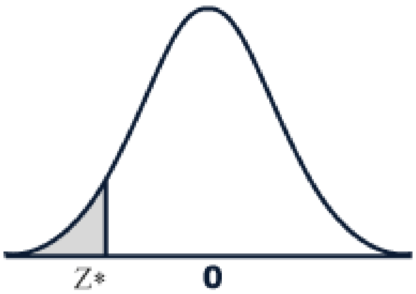
Slide 12:

The P-Value for a One-Sided Test

The p-value represents the probability of obtaining the calculated test statistic, or a value that is more extreme. A more extreme value would be any value that exceeds the test statistic in the same direction as the critical region. It represents the area delimited by the test statistic that contains the critical region.

For example, if we are looking at a one-sided test to the right (Ha: $\mu > ?$), we would calculate the area that is contained immediately to the right of the test statistic. If the situation was a one-sided test to the left (Ha: $\mu < ?$), the area immediately to the left of the test statistic would be used. To find the area, **use the normal distribution table, regardless of the test statistic** (z^* or t^*). Although this is no problem for an already-established z-test situation, this procedure will likely incur some error for a t-test. In case of a conflict with the conclusions of the graphical method and the p-value when using t^*, the graphical method takes precedence since its values came directly from the t-table. This is a rare occurrence, however, and would mean that the test statistic is very close to the critical region.

One-Sided Tests to the Left



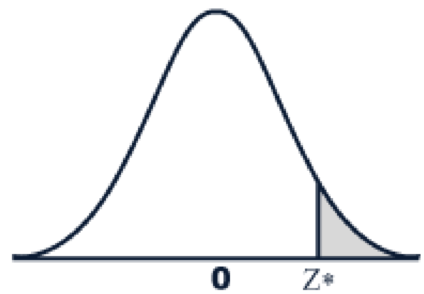
Ho: μ ≥ 0
Ha: μ < 0

The p-value is represented by the area that is located immediately to the left of the test statistic: **P(Z < Z*)**

One-Sided Tests to the Right

Ho: μ ≤ 0
Ha: μ > 0

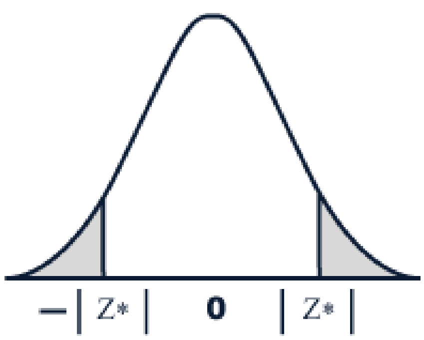
The p-value is represented by the area to the right of the test statistic: **P(Z > Z*)**



Slide 13:

The P-Value for a Two-Sided Test

For a two-sided test, we start at the test statistic and go in the direction of the closest critical region. That is, if our test statistic is positive, we go to the right. If it is negative, we go to the left. Then, since the critical area is located at both extremities, we multiply the p-value by 2.



Ho: μ = 0
Ha: μ ≠ 0

The p-value is represented by twice the area to the left of a negative test statistic, or twice the area to the right of a positive one.

2 [P(Z > | Z* |)]

Once you have obtained your p-value, you compare it to the alpha value (significance level). A small p-value (less than α) leads to the rejection of Ho, whereas a large p-value (greater than α) does not.

In our case, we have a two-sided test with the test-statistic being located at -2.37. This means that we will use the area found to the left of z*, and then multiply it by 2 to get our p-value.

P(z < -2.37) = 0.0089
2 x P(z < -2.37) = 2 x 0.0089 = **>0.0178**

Since this value (0.0178) is less than our alpha value (0.05), we have enough evidence to reject

the null hypothesis and accept the alternate one. In other words, the hospital's claim is false. Therefore, the average time it takes the hospital to receive the patients in the emergency ward is not 20 minutes. Completing a hypothesis test using only the p-value allows us to skip over step 3.

So, what can we conclude about this test? Both the classical method as well as the p-value point to the fact that the null hypothesis must be rejected. This means that the hospital's claim is false. However, can we say that there is enough evidence to support the claim that the time of arrival is less than 20 minutes? Although the test statistic seems to indicate this (since the sample mean is 18.5 minutes, and the test statistic is in the critical region), we cannot make this assumption unless we perform a new, one-sided test. Only in doing so can we have the proof needed to support the new claim.

Slide 14:

Type I and Type II Error

As is the case with all statistics, there is always room for error. It is quite possible that the conclusion we reached in our hypothesis test is false. Be it because of faulty calculations, data gathering methods, or a random occurrence, it is possible that a mistake was made. The possible mistakes are as follows:

	Accept Ho	**Reject Ho**
Ho is true	**Correct Decision**	Type I Error
Ho is false	Type II Error	**Correct Decision**

If the null hypothesis (Ho) is true and it is accepted, or false and rejected, the decision is, in either case, a correct one. However, if the null hypothesis Ho is true but it is rejected, this is an error. Accepting the alternate hypothesis (Ha) when it should have been rejected is also a possible error.

The first cause for error, called **Type I error**, is the probability of rejecting Ho when you should not have rejected it. The probability of committing such an error is denoted by the Greek letter α (alpha). The second source for error is called a **Type II error**. This represents the probability of accepting Ho when you should not have done so. The probability of committing a type II error is described by the Greek letter ß (beta), which is also derived from the significance level.

Since this is an introductory statistics course, we will not concern ourselves with the calculation of each type of error. The concept of hypothesis test errors, and more specifically, the power of the statistical tests, are issues covered in more advanced statistics courses. For more information on these concepts, check out the following website: >www.math.com.

Slide 15:

Hypothesis Test

Example 1

A race car driver claims that his vehicle can attain 250 Km/h in 15 seconds. Can you test his claim at a 95% confidence level? A random sample of ten trials was taken and these were the results:

252	246	242	250	255	258	250	252	250	258

From the data, we can determine the sample mean (251.3), the sample standard deviation (4.99), as well as the degrees of freedom (10-1=9). Since not only is the sample size small, but the population standard deviation is also unknown, we must use the t-statistic. Since this is a two-sided test, we must divide alpha into two parts (a = 1 - 0.95 = 0.05; a/2 = 0.05/2 = 0.025).

Therefore, the CI would be:

$$\bar{x} - t_{(n-1,\alpha/2)}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{(n-1,\alpha/2)}\frac{s}{\sqrt{n}}$$

The lower limit = 251.3 - $t_{(9,\ 0.025)}$ x (4.99/sqrt 10) = **247.73**

The upper limit = 251.3 + $t_{(9,\ 0.025)}$ x (4.99/sqrt 10) = **254.87**

*Note that the $t_{(9,\ 0.025)}$ requires the use of the t-distribution table where degrees of freedom = 9, and 0.025 represents the alpha level. Therefore, **$t_{(9,\ 0.025)}$ = 2.262**.

The confidence interval would indicate that we do not reject the null hypothesis since it is highly probable that 250 km/h lies within this interval (95% confidence). There is, however, a chance of making a type II error.

Slide 16:

Hypothesis Test (Cont'd)

Example 2

A sample of 36 measurements was taken to test the hypothesis that the mean level of carbon monoxide (CO) in downtown Montreal air is greater than 4.9 parts per million. Given a population standard deviation of 1.8, and a sample mean of 5.3, test this hypothesis at the 95% confidence level.

***Step 1: Determine the hypothesis:**

Ha: μ > 4.9 ppm, so, consequently, Ho: μ ≤ 4.9 ppm
(Since we want to focus on the fact that the CO is greater than 4.9).

***Step 2: Determine the level of significance:**

α = 0.05

***Step 3: Determine the critical region(s):**

Z(α) for 95% is known to be 1.645 (one-sided). This value came from the normal distribution table and it marks the point on the graph where there is an area of 0.95 to the left of the Z value, and 0.05 (the critical region) on the right. Therefore, our boundary for the critical region is 1.645. Any test statistic greater than this value will cause the rejection of the null hypothesis since it would mean that the test statistic falls within the critical region.

***Step 4: Determine the test statistic:**

Z* = ($\bar{x}$ - μ) / (s / sqrt n) = (5.3 - 4.9) / (1.8 / sqrt 36)
Z* = 1.333

***Step 5: Conclusion**

Since 1.333 does not fall into the critical region, there is not enough evidence to reject Ho. Therefore, there is not enough evidence to conclude that the mean level of CO in Montreal's atmosphere is greater than 4.9 ppm.

Slide 17:

Hypothesis Test (Cont'd)

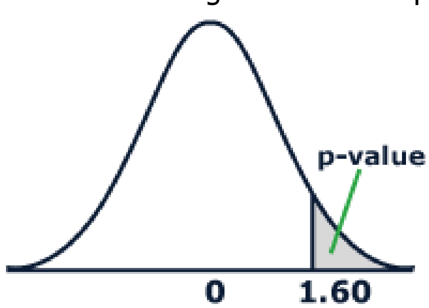
Example 3

Patrick has recorded the grades of all students taking INTE 296 since its inception in the winter of 1996. The mean of these grades is μ = 72. The current class of 36 students seems to be better than average, and the instructor wants to show that, according to their average, the current class is superior to the previous classes. Does the current class average of 75.2 present enough evidence to support his bold claims? Use an alpha of 0.05 and a standard deviation of 12.

First of all, we must decide on our hypotheses:
Ho: μ ≤ 72 (class is not superior)
Ha: μ > 72 (class is superior)

Using the Z-score formula (since the sample size is large), we calculate the Z* as being **1.60**.

By drawing a rough sketch of the standard curve, we are able to plot Z*=1.60 on our graph, and determine, with the aid of the Z-score table, the area representing the values that are greater than Z* (p-value). In this case, the area to the right of 1.60 is equal to:



P(z > z*) = P(z > 1.60) = 1-P(Z < 1.60) = 1-0.9452 = 0.0548.

Since the p-value is greater than our significance level (0.05), we fail to reject Ho. Note, however, that had the significance level been 0.10, we would have rejected Ho in favour of Ha! HaHaHa!

Slide 18:

Video Examples

Introduction to Hypothesis Testing

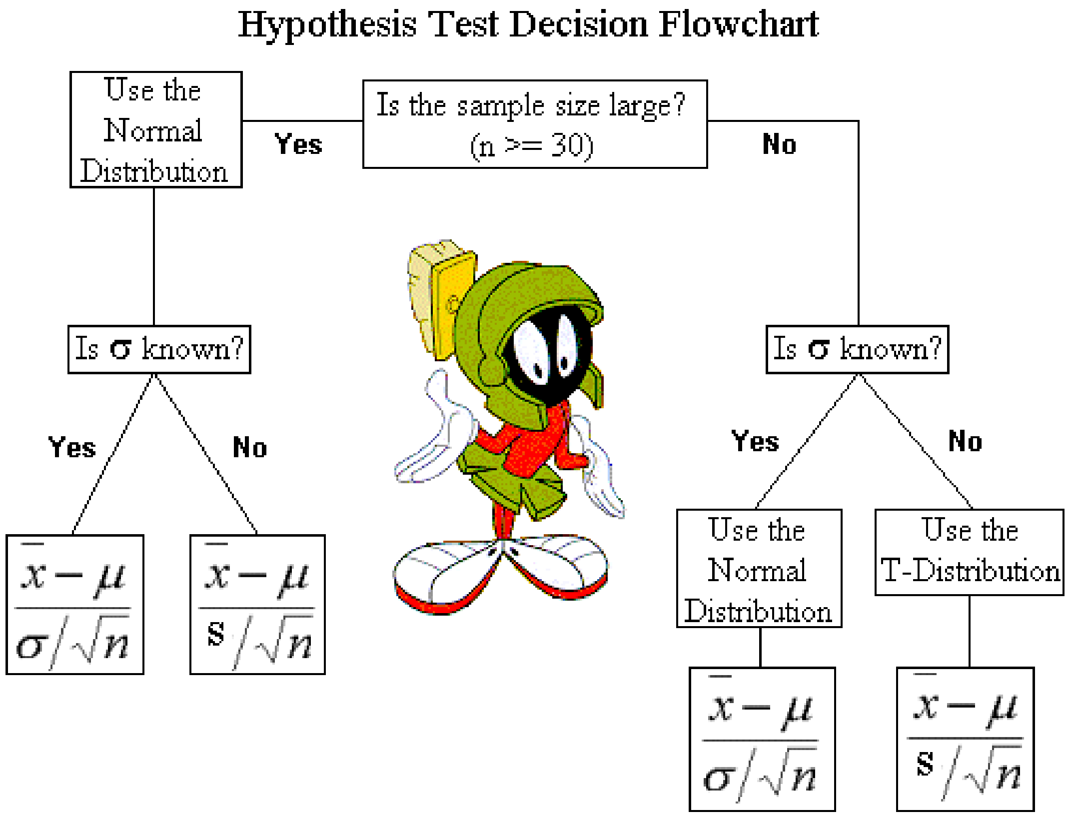
These introduction videos provide additional explanations on hypothesis testing.

- View streaming video if you have a high-speed Internet connection.
- Download video if you have a low-speed Internet connection.

1. Setting up the hypotheses	Streaming video (Length 05:10) Download video (97.1 MB)
2. Critical region	Streaming video (Length 05:58) Download video (130 MB)
3. Test statistic	Streaming video (Length 03:58) Download video (77.9 MB)
4. P-value	Streaming video (Length 13:40) Download video (177 MB)
5. Conclusion	Streaming video (Length 02:57) Download video (28.7 MB)

Slide 19:

Hypothesis Test Decision Flowchart



Slide 20:

Recap

Hypothesis testing is quite procedural, as it will always follow a series of sequential steps. It is important to start things off on the right foot and correctly set up the null and alternate hypothesis before carrying out the test using one of three methods. In this lesson, you also learned that:

- The claim made in a hypothesis test does not necessarily have to be represented by the null hypothesis.
- The null hypothesis (Ho) must always have an equal sign in it, be it =, <=, or >=, and the Ha is the opposite.
- The confidence interval method for conducting a hypothesis test can only be used in a two-sided situation.
- The p-value is the probability that another sample of the same size will yield a test statistic that is equal to, or more extreme than, the current one.
- If the test statistic falls in the critical region, and the p-value is less than the significance level, the null hypothesis is rejected in favour of the alternate one.
- It is very rare that the population variance is given (or known). Therefore, the best

estimate for it is the sample standard deviation value.

You can post a message online in your discussion folder any time you have something to share with your discussion group concerning the current lesson. Simply click Discussion Board or use the menu at the top of the screen.

Next lesson: Dependent Means Hypothesis Testing